

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
13 January 2005 (13.01.2005)

PCT

(10) International Publication Number  
**WO 2005/003920 A2**

(51) International Patent Classification<sup>7</sup>: **G06F**

(21) International Application Number: **PCT/US2004/021294**

(22) International Filing Date: 1 July 2004 (01.07.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/484,818 2 July 2003 (02.07.2003) US  
10/645,313 20 August 2003 (20.08.2003) US

(71) Applicant: **VIBRANT MEDIA [US/US]**; 19823 58th Place, South, Kent, WA 98032 (US).

(72) Inventors: **STEVENSON, Douglas, B.**; 4238 23rd Street, San Francisco, CA 94114 (US). **GOODING, Craig, Allen**; 30 Alma Street, London, GB NW5 (US).

(74) Agent: **GRAHAM, Lawrence, D.**; Black Lowe & Graham PLLC, 701 Fifth Avenue, Suite 4800, Seattle, WA 98104 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SI, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**  
— without international search report and to be republished upon receipt of that report

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

WO 2005/003920 A2

(54) Title: METHOD AND SYSTEM FOR AUGMENTING WEB CONTENT

(57) Abstract: A system for augmenting data from a source data file with data from a reference database to generate an augmented data file is provided. The system includes a reference database including at least one reference datum. A handler component is configured to retrieve a source data file including the structured datum. A locator component is configured to locate the structured datum in the source data file; an analyzer component configured to associate the identified structured datum to one reference datum to create an association according to an analyzing strategy. A generating component is configured to generate a hyperlink based upon the association and embeds the generated hyperlink in the source file to create an augmented data file.

**METHOD AND SYSTEM FOR AUGMENTING WEB CONTENT**

10

INVENTORS  
Douglas Stevenson  
Craig Allen Gooding

**BACKGROUND OF THE INVENTION**

Hypertext is the organization of computer-based text into connected associations enabling a user to quickly access information that the user chooses. An instance of such an association is called a hyperlink or hypertext link. Hypertext was the main concept that led to the invention of the World Wide Web, which is nothing more than an enormous amount of information content connected by an enormous number of hyperlinks.

While the hyperlink has proven to be a successful means of relating two pieces of information, the process of generating hyperlinks has proven to be generally tedious. To create a single link, the process requires an author to define such portions, usually a text string or photo element of a structured file (usually a text document, web page, or other form of document) from which the hyperlink originates and a destination address at which the hyperlink terminates. In a closed system such as in a local network of workstations, the destination might be within the same file, directory, or computer, or the destination may be a designated file within a designated directory on the network.

Even with the required information, the knitting together of hyperlinks still requires some skill. The would-be author of a document with suitable links must first identify the content of the file the author seeks to augment and then must use an appropriate application to edit the file. Generally, the MIME header embedded in the file identifies the file type. The embedded header 5 allows a computer software product to recognize the data by virtue of its Multi-Purpose Internet Mail Extensions (“MIME”) type. MIME is an extension of the original Internet e-mail protocol that lets people use the protocol to exchange different kinds of data files on the Internet including audio, video, images, application programs, and other kinds, including text generally in the ASCII format. Once identified, the file is opened for review using the appropriate application as 10 identified by the MIME header.

When the network is broader, such as on the Internet, static addresses on the Internet may be used as destinations. A Uniform Resource Locator is the address of a file (resource) accessible on the Internet. The type of resource depends on the Internet application protocol. Using the World Wide Web’s protocol, Hypertext Transfer Protocol (HTTP), the resource can be 15 an HTML page, an image file, a program such as a common gateway interface application or Java applet, or any other file type supported by HTTP. The URL contains the name of the protocol required to access the resource, a domain name that identifies a specific computer on the Internet, and presents a hierarchical description of a file location on the computer.

For this reason, the content of resources may not have all of the links that would be 20 useful. Old data might direct browsers to addresses where no data is now stored. “Link rot,” describes a gradual loss of data at URL’s linked to documents. This gradual loss occurs when a destination document is removed while the link in an originating document to the destination document remains. The reader receives a “404 message,” an arbitrarily assigned code indicating that the page to which the reader has directed the browser no longer exists at the designated 25 address. Another form of link rot occurs when the destination page has been changed in content and is no longer relevant according to the sending description.

Where parties, such as advertisers, wish to inject links into existing resources in order to direct the reader's browser to designated resources, it is critical that the links remain current. Because files are static, links that are old will not complete the hyperlink transit and therefore will lose the benefit of the hyperlink. Fixedly embedding links in the file subjects the file to link 5 rot.

There is, therefore, an unmet need in the art for a publishing system and method for augmenting resources and maintaining suitable current hyperlinks within the resources.

#### SUMMARY OF THE INVENTION

A system for augmenting data from a source data file with data from a reference database 10 to generate an augmented data file is provided. The system includes a reference database including at least one reference datum. A handler component is configured to retrieve a source data file including the structured datum. A locator component is configured to locate the structured datum in the source data file; an analyzer component configured to associate the identified structured datum to one reference datum to create an association according to an 15 analyzing strategy. A generating component is configured to generate a hyperlink based upon the association and embeds the generated hyperlink in the source file to create an augmented data file.

One presently preferred embodiment includes a system for augmenting file content, including web content, with hyperlinks to designated destinations. The system works based on 20 finding a datum (a data subset of a file) within a file, recognizing the datum based upon the contents of a reference database, associating the datum with a designated resource (in a presently preferred embodiment by means of a uniform resource locator address) and generating a hyperlink in the data source file.

The generated hyperlink (in a presently preferred embodiment) receives a user-friendly 25 name based on the contents of the resource located at the uniform resource locator address. The

an embodiment is a add-on to a browser allowing the browser to augment files "on the fly," i.e. where the user directs the browser to a resource located on a network, the method analyzes the file as it is opened by the browser, augments the file with appropriate hyperlinks, and displays the augmented file with active hyperlinks. "Clicking on" the hyperlink will redirect the browser to 5 the associated uniform resource locator address.

Another presently preferred embodiment provides a rigorous procedure for augmenting files that assures that a greater number of hyperlinks are more uniformly applied than is the rule with human authoring of hyperlinks. A reference database can check the content of the file and will always place a hyperlink where appropriate based upon context. Because the reference 10 database can be readily updated, the invention assures that redirection to current resources, preventing the dead link, i.e. "Error 404, file not found."

A presently preferred embodiment provides a method and a software product to add advertisements to existing web content by hyperlinking occurrences of structured data such as text strings to resources located at a uniform resource locator address.

15 As will be readily appreciated from the foregoing summary, the invention provides a system and a method for rapidly and for rigorously augmenting files with hyperlinks.

#### **BRIEF DESCRIPTION OF THE DRAWINGS**

The preferred and alternative embodiments of the present invention are described in detail below with reference to the following drawings.

20 FIGURE 1 is a flowchart of the method for augmenting source data files;  
FIGURE 2 is a block diagram of a system for augmenting source data files;  
FIGURE 3 is screen shot of a source data file selected for augmentation;  
FIGURE 4 is a block of text selected for augmentation;  
FIGURE 5 is the block of text selected for augmentation based upon the occurrence of a  
25 text string;

FIGURE 6 is an excerpt of an exemplary reference database;

FIGURE 7 is a screen shot of the source data file selected for augmentation showing hyperlink and associated user-friendly name;

FIGURE 8 is a screen shot of the contents of a destination file upon redirection of the  
5 browser.

#### **DETAILED DESCRIPTION OF THE INVENTION**

By way of overview, a system for augmenting data from a source data file with data from a reference database to generate an augmented data file is provided. The system includes a reference database including at least one reference datum. A handler component is configured to 10 retrieve a source data file including the structured datum. A locator component is configured to locate the structured datum in the source data file; an analyzer component configured to associate the identified structured datum to one reference datum to create an association according to an analyzing strategy. A generating component is configured to generate a hyperlink based upon the association and embeds the generated hyperlink in the source file to create an augmented data 15 file.

FIGURE 1 is a flowchart of the method 9 for augmenting source data files. At an appropriate starting terminus 10, the method begins by reading a structured datum from a source data file at a block 13. The source data file may be one designated by an input uniform resource locator address or by any suitable means to designate a resource. Upon opening, the method 9 20 may optionally identify the type of content on the page with a content identifier such as a MIME header. In one embodiment of the invention, the method 9 merely searches for the presence of a reference datum, either informed by the content identifier or by simply searching an occurrence of a well-structured datum within a given file. However, once the file is open, the method has the contents of the file available for comparison to a reference database.

At a block 16, the method 9 locates an occurrence of a reference datum corresponding to the structured datum read in the source data file. One presently preferred means of discerning the correspondent relation between the structured datum and the reference datum is by a JavaScript call made to a web-enabled database. The Java script then compares the contents of the source data file with reference data stored in a web-enabled reference database. In one presently preferred embodiment, the reference and structured data are keywords. The JavaScript code then extracts the text from the document and converts all of the keywords in the document to hyperlinks.

Other analyzing strategies are also available. Though keywords are a facile and efficient means of generating hyperlinks. One presently preferred embodiment uses a “fuzzy expert” or a neural network analysis of the source data file, such as by a natural language search of the document to generate a distinct identifier for the content in the source data file. One advantage of a natural language search is the ability to better place content in context making links more contextually appropriate, for instance, security might relate to security of a physical plant such as security of a residence in one source data file in one context and security of a website in another. Natural language searches, however, create a large processing overhead, making them less desirable where such resources are at a premium.

At a block 19, the method 9 generates an association based upon the located reference datum in the reference database. The reference datum will include not only the identifier, such as a keyword in one embodiment, but also the associated uniform resource locator address of the intended destination based upon the occurrence of the identifier in the source data file. Generating an association means to hyperlink the located structured datum in the source data file to the associated uniform resource locator as found in the reference datum in the reference database. The generated hyperlink might optionally include a user-friendly description of the content of the resource found at the associated uniform resource locator address and additionally might include an additional identification code such as an “advertiser id.” In each embodiment,

the generated hyperlink is added to the original source data file at a block 22 resulting in an augmented data file. Having generated the augmented data file, the method 9 then terminates at a block 25.

FIGURE 2 is a block diagram of a system for augmenting source data files. In this 5 exemplary system, a network 33 is shown in a presently preferred embodiment. Those skilled in the relevant art will readily appreciate that the system may be practiced without the presence of a network link. Also a reference database 39 is shown as directly connected to the locator 42 and the analyzer 45. The system is not compromised by network links to any of the several components. One presently preferred embodiment has the reference database 39 set on a 10 web-enabled page for remote calls through the Internet to the database. Like the presence of the network 33 described above, the absence of a defined link through the Internet does not compromise the operation of the method.

A source data file 30 resides on a server on a network 33. A handler 36 retrieves the source data file 30 for use by the system. A locator 42 examines the retrieved source data file 30 15 for comparison to the reference database 39 according to an analyzing strategy. The locator 42 designates found structured data from the source data file 30 and found reference data from the reference database 39 and provides the reference data to an analyzer 45.

The analyzer 45 is used to create associations between each found structured datum and the uniform resource locator address within the corresponding reference datum found by the 20 locator 42. These associations define the nature of a hyperlink a generator 48 generates according to the association created at the analyzer 45. The generator 48 embeds these hyperlinks in the source data file 30. The resulting augmented data file 50 is returned to the handler 36 to reside at a uniform resource locator address on the network 33.

FIGURE 3 is a screen shot 101 of contents of a source data file 30 (FIGURE 2) selected 25 for augmentation. For the purposes of this exemplary discussion, a block of text 104 is selected by the method for augmentation. The present invention does not require the opening of the

source data file 30 in a browser however for the purposes of illustration, the screen shot 101 is provided herein.

FIGURE 4 is a block of text 107 selected for augmentation. The locator 42 (FIGURE 2) begins its analysis of the selected block by isolating the block of text 107 as structured data.

5 FIGURE 5 is the block of text 107 selected for augmentation indicating the occurrence of a text string 111. The locator 42 isolates the occurrence of the text string 111 because of its presence in the reference database 39.

Referring to FIGURES 2 and 6, while the excerpt 115 of the exemplary reference database 39 shows it to be a “flat file” database, any database 39, relational, flat file, or other 10 configuration will suitably fulfill the basic functions of associating an identifier such as a text string 118, with a uniform resource locator address 124, and optionally a user-friendly description 121 of the contents of the file found at the associated uniform resource locator address 124. (Not shown is the optional “advertiser id.”)

The locator 42 (FIGURE 2) refers to the database as it reviews the contents of the source 15 data file 30 (FIGURE 2). According to the analyzing strategy, the text string 118 occurs at a reference datum in the source data file 30, the locator 42 provides the uniform resource locator 124 associated with the found text string 118 along with the occurrence of the text string in the reference datum to the analyzer 45. Upon receiving the ordered pair from the locator address 42, the analyzer 45 creates the association. With the association from the 20 analyzer 45, the generator 48 creates a hyperlink. As a result, the text block receives the appropriate hyperlink.

FIGURE 7 is a screen shot 101 of the augmented data file 50 showing the selected block of text 104, hyperlink 129, and associated user-friendly name 133. As will readily be appreciated by those skilled in the relevant art, the augmented data file 50 need not reside on a network. For 25 instance, where a browser is enhanced with the inventive method 9 (FIGURE 1), the browser might be directed to content on the Internet. Upon opening the source data file 30 (FIGURE 2),

the browser will effect the method 9 (FIGURE 1) such that rather than displaying the contents of the source data file 30, the browser will display the augmented data file 50. The resulting augmented files 50 would be current at the time of viewing. In a publisher embodiment, the advertiser might continually files on the network, replacing augmented files on the network.

5        The augmented data file 50 is displayed as set forth in FIGURE 7. When a reader places the mouse cursor over the hyperlink 129, a new layer for display showing the associated user-friendly name 133 may be optionally evoked. The “advertiser id” may be optionally embedded but not necessarily visible, the advertiser id. This advertiser id specifically provides a means for tracking the number of times the hyperlink 129 is activated and generating a record for  
10      tracking revenue due to advertising. The reader clicks on the hyperlink 129 to direct the browser to a destination file.

FIGURE 8 includes a screen shot 140 of the browser displaying the contents of the destination file located at the uniform resource locator address 137 upon redirection of the browser. When the reader clicks on the hyperlink 129, a new browser window is opened and  
15      directed through a click-tracking server to the hyperlink destination.

While the preferred embodiment of the invention has been illustrated and described, as noted above, many changes can be made without departing from the spirit and scope of the invention. For example, the method 9 will generate static files on the network at distinct uniform resource locator addresses 124 and 137 in order to distinguish from the original source data files.  
20      Accordingly, the scope of the invention is not limited by the disclosure of the preferred embodiment. Instead, the invention should be determined entirely by reference to the claims that follow.

What is claimed is:

1. A system for augmenting data from a source data file with data from a reference database to generate an augmented data file, comprising:
  - a reference database including at least one reference datum;
  - 5 a handler component configured to retrieve a source data file including at least one structured datum;
  - a locator component configured to locate the structured datum in the source data file;
  - an analyzer component configured to associate the identified structured datum to the one reference datum to create an association according to an analyzing strategy;
- 10 and
  - a generating component configured to:
    - generate a hyperlink based upon the association; and
    - embed the generated hyperlink in the source data file to create an augmented data file.
- 15 2. The system of Claim 1, wherein the source data file is stored at an address on a network.
3. The system of Claim 2, wherein the network is a public network.
4. The system of Claim 3, wherein the network is the Internet.
5. The system of Claim 1, wherein the structured datum includes a formatted webpage.
- 20 6. The system of Claim 1, wherein the reference datum includes a first uniform resource locator address.
7. The system as defined in Claim 7, wherein the first uniform resource locator address is the location of an advertisement.

8. The system of Claim 7, wherein the reference datum includes generating a second identifier associated with the first uniform resource locator address.
9. The system of Claim 8, wherein the analyzer component is further configured:
  - 5 to generate a first identifier based upon the structured datum;
  - to locate the second identifier corresponding to the first identifier; and
  - to associate the first uniform resource locator address with the structured datum.
10. The system of Claim 9, wherein the first uniform resource locator address is further associated with a user-friendly descriptor.
11. The system of Claim 9, wherein the first identifier is a first text string.
- 10 12. The system of Claim 11, wherein the first text string includes keywords.
13. The system of Claim 12, wherein the second identifier is a second text string.
14. The system of Claim 13, wherein the analyzer component is further configured to locate the second text string according to the first text string.
15. The system of Claim 14, wherein the analyzer component is further configured to associate the first uniform resource locator address with the first text string.
16. The system of Claim 9, wherein the analyzer component generates the first identifier by means of a “fuzzy expert” search engine.
17. The system of Claim 16, wherein the analyzer component generates the first identifier by means of a natural language search engine.
- 20 18. The system of Claim 17, wherein the hypertext link is further associated with a user-friendly descriptor.

19. The system of Claim 1, wherein a publishing component is configured to save the augmented file at a desired second uniform resource locator address.
20. The system of Claim 1, wherein a browser is configured to display the augmented data file when directed to the source data file.
- 5 21. A method of augmenting data from a source data file with data from a reference database, the method comprising:
  - retrieving at least one data file including at least one structured datum from a first address;
  - identifying the structured datum;
  - 10 locating reference datum from a reference database according to the identified structured datum;
  - generating a hyperlink according to the reference datum;
  - augmenting the data file with the hyperlink; and
  - storing the augmented data file at a second address.
- 15 22. The method of Claim 21, wherein the first address is on a network.
23. The method of Claim 22, wherein the network includes at least a portion of the Internet.
24. The method of Claim 22, wherein retrieving one data file includes retrieving a web page.
25. The method of Claim 21, wherein locating reference datum includes locating a uniform 20 resource locator address.
26. The method of Claim 25, wherein generating a hyperlink includes associating the structured datum to a uniform resource locator address.

27. The method of Claim 26, wherein generating a hyperlink includes generating a user-friendly description of the data contained at the uniform locator address.
28. The method of Claim 21, further comprising displaying the augmented data file.
29. The method of Claim 28, wherein displaying the augmented data file includes displaying the user-friendly name.  
5
30. A method of augmenting structured data stored in a source data file with unstructured data stored in a reference database, comprising:
  - reading a structured datum from a source data file;
  - locating a reference datum in a reference database corresponding to the read structured datum;
  - 10 generating an association to the read structured datum based upon the located reference datum; and
  - augmenting the source data file with the generated association.
31. The method of Claim 30, wherein the source data file is stored on a network.
- 15 32. The method of Claim 31, wherein the source data file is stored on the Internet.
33. The method of Claim 30, wherein locating the reference datum includes locating a uniform resource locator address.
34. The method of Claim 33, wherein locating the uniform resource locator address includes locating the uniform resource locator address for an advertisement.
- 20 35. The method of Claim 30, wherein locating a reference datum in a reference database corresponding to the read structured datum and wherein corresponding includes corresponding according to an analyzing strategy.

36. The method of Claim 35, wherein the analyzing strategy is locating a first text string in the structured datum and matching a second text string in the reference datum.
37. The method of Claim 35, wherein the analyzing strategy comprises locating a first keyword in the structured datum to correspond to a second key word in the reference datum.
- 5 38. The method of Claim 35, wherein the analyzing strategy includes generating a first identifier to the structured datum and locating a second identifier in the reference datum matching the first identifier.
39. The method of Claim 38, wherein the generating a first identifier is based upon a “fuzzy expert” search engine.
- 10 40. The method of Claim 39, wherein generating the first identifier further includes retrieving the natural language text from the structured datum.
41. A system for associating data in a reference database with structured data in a source data file, comprising:
  - 15 means for reading a structured datum from the source database;
  - means for locating a reference datum in a reference database corresponding to the read structured datum;
  - means for generating an association to the read structured datum based upon the located reference datum; and
- 20 means for augmenting the source data file with the generated association.
42. The system of Claim 41, wherein the source data file is stored on a network.
43. The system of Claim 41, wherein the source data file is stored on the Internet.

44. The system of Claim 43, wherein locating the reference datum includes locating a uniform resource locator address.
45. The method of Claim 44, wherein locating the uniform resource locator address includes locating the uniform resource locator address for an advertisement.
- 5 46. The method of Claim 41, wherein locating a reference datum in a reference database corresponds to the read structured datum, and wherein corresponds includes corresponding according to an analyzing strategy.
- 10 47. The method of Claim 41, wherein the analyzing strategy includes locating a first text string in the structured datum and locating a second text string in the reference datum matching the first text string.
48. The method of Claim 41, wherein the analyzing strategy includes matching a first keyword in the structured datum to second key word in the reference datum.
49. The method of Claim 41, wherein the analyzing strategy includes generating a first identifier to the structured datum and matching a second identifier in the reference datum.
- 15 50. The method of Claim 49, wherein the generating a first identifier is based upon a “fuzzy expert” search engine.
51. The method of Claim 50, wherein generating the first identifier further includes retrieving the natural language text from the structured datum.
- 20 52. A computer software program stored on a readable medium, the computer software program comprising:  
means for reading a structured datum from the source database;

means for locating a reference datum in a reference database corresponding to the structured datum;

means for generating an association to the structured datum based upon the reference datum; and

5 means for augmenting the source data file with the association.

53. The software program of Claim 52, wherein the source data file is stored on a network.

54. The software program of Claim 53, wherein the source data file is stored on the Internet.

55. The software program of Claim 53, wherein locating the reference datum includes

10 locating a uniform resource locator address.

56. The software program of Claim 55, wherein locating the uniform resource locator address includes locating the uniform resource locator address for an advertisement.

57. The software program of Claim 52, wherein locating a reference datum in a reference database includes locating a reference datum corresponding to the read according to

15 an analyzing strategy.

58. The software program of Claim 52, wherein the analyzing strategy is locating a first text string in the structured datum and matching a second text string in the reference datum.

59. The software program of Claim 52, wherein the analyzing strategy comprises matching

20 a first keyword in the structured datum to second key word in the reference datum.

60. The software program of Claim 52, wherein the analyzing strategy is generating a first identifier to the structured datum and matching a second identifier in the reference datum.
61. The software program of Claim 60, wherein generating a first identifier is based upon a 5 “fuzzy expert” search engine.
62. The software program of Claim 60, wherein generating the first identifier further includes retrieving the natural language text from the structured datum.
63. A computer software program stored on a readable medium for augmenting data from a source data file with data from a reference database to generate an augmented data 10 file, comprising:
  - a reference database including at least one reference datum;
  - a handler component configured to retrieve a source data file including at least one structured datum;
  - a locator component configured to locate the structured datum in the source data file;
  - 15 an analyzer component configured to associate the identified structured datum to one reference datum to create an association; and
  - a generating component configured to:
    - generate a hyperlink based upon the association; and
    - embed the generated hyperlink in the source file to create an augmented data 20 file.
64. The software program of Claim 63, wherein the source data file is stored at an address on a network.
65. The software program of Claim 64, wherein the network is a public network.

66. The system of Claim 64, wherein the network is the Internet.
67. The system of Claim 64, wherein the structured datum includes a formatted webpage.
68. The system of Claim 63, wherein the reference datum includes a first uniform resource locator address.
- 5 69. The system as of Claim 68, wherein the first uniform resource locator address is the location of an advertisement.
70. The system of Claim 68, wherein the reference datum includes a second identifier associated with the first uniform resource locator address.
- 10 71. The system of Claim 68, wherein the analyzer component is further configured to: generate a first identifier based upon the structured datum; to locate the second identifier corresponding to the first identifier; and to associate the first uniform resource locator address with the structured datum.
72. The system of Claim 70, wherein the first uniform resource locator address is further associated with a user-friendly descriptor.
- 15 73. The system of Claim 71, wherein the first identifier is a first text string.
74. The system of Claim 73, wherein the first text string includes keywords.
75. The system of Claim 74, wherein the second identifier is a second text string.
76. The system of Claim 71, wherein corresponding includes corresponding based upon an analyzing strategy.

77. The system of Claim 76, wherein the analyzing strategy includes matching the first text string with the second text string and further associating the first uniform resource locator address with the first text string.
78. The system of Claim 77, wherein the analyzing strategy generates the first identifier by means of a “fuzzy expert” search engine.  
5
79. The system of Claim 77, wherein the analyzing strategy generates the first identifier by means of a natural language search engine.
80. The system of Claim 70, wherein the hypertext link is associated with a user-friendly descriptor, the user-friendly descriptor being associated with the associated first uniform resource locator address.  
10
81. The system of Claim 63, wherein a publishing component is configured to save the augmented file at a desired second uniform resource locator address.
82. The system of Claim 63, wherein a browser is configured display the augmented data file when directed to the source data file.

1/8

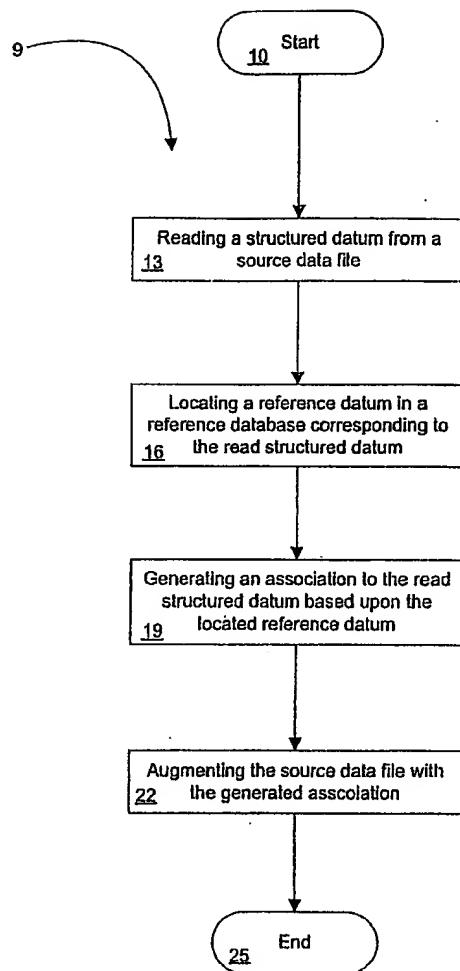


FIG. 1

2/8

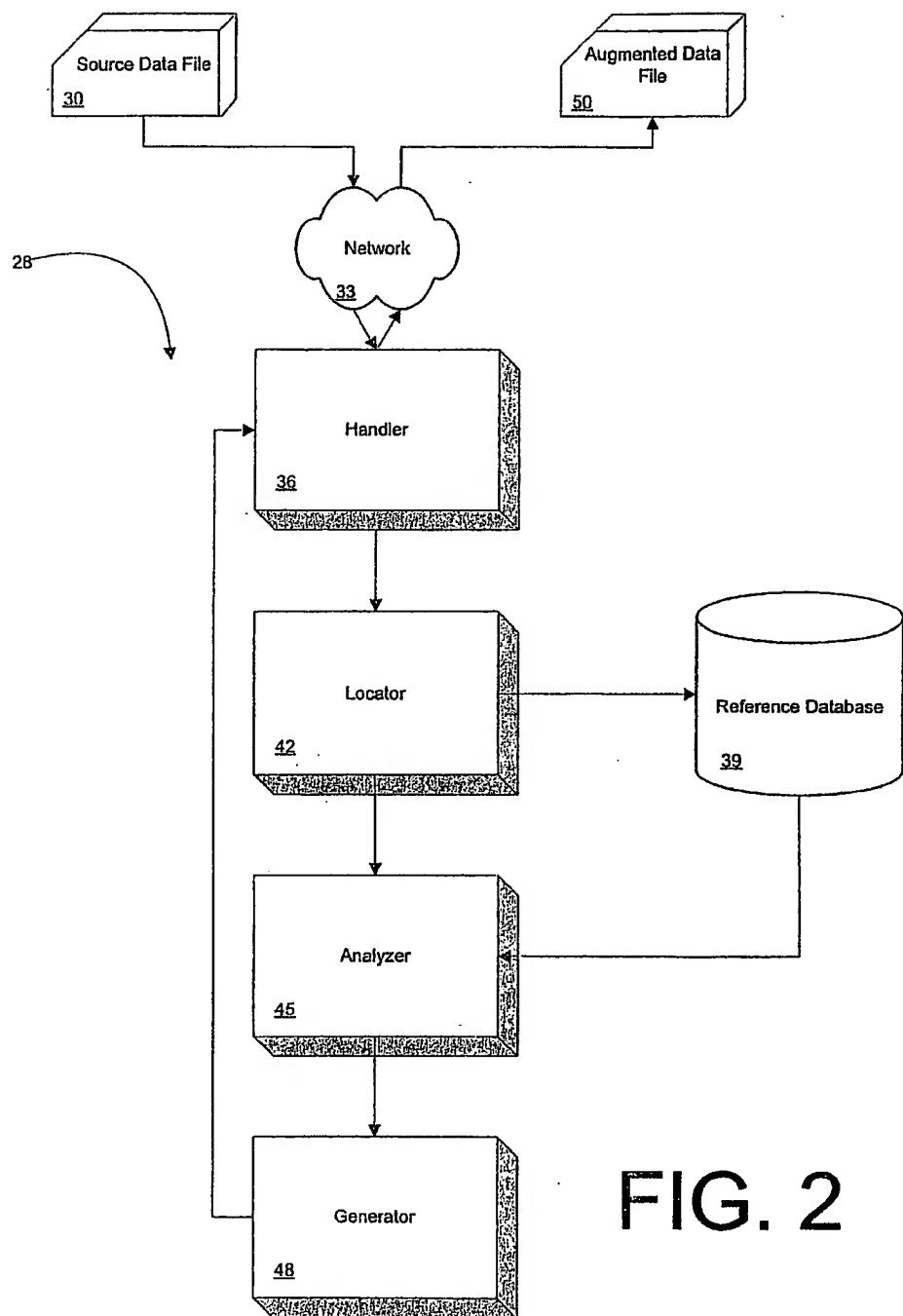


FIG. 2

3/8

File Edit View Favorites Tools Help

Back Search Favorites Back Forward Stop Refresh Home

Address: http://www.computing.net/security/www/board/forums/1019.html

Please let me know - driving me crazy!

My Computing Net

Online Section

Site Search

Desktop Search

Message Board

FAQs

Install Drivers

About

Thanks

Home

Thanks, Brian

Response Number 1

Name: Brian

Date: April 22, 2003 at 04:21:27 Pacific

Subject: Virus Help!!!!

Reply:

Sounds like you stopped it dead in its tracks. You could do a second virus scan at HOUSE CALL at the Trend Micro website if that would ease your mind. Brian, if you are going to use KAZAA, remember to be very cautious and scan everything. As many people that try to help people at the peercy area know fid1 well, there are virus/trojans/worms that are only available by using KAZAA!

104

Please let me know - driving me crazy!

Response Number 2

Name: Brian

Date: April 22, 2003 at 05:45:15 Pacific

Subject: Virus Help!!!!

Reply:

Thanks for your advice capt I am as careful as i can be on the catastrophe that is kazaa. And at least i know my av is working hah? lol.

Thanks again for your time

Brian

Response Number 3

Name: Brian

Date: April 22, 2003 at 05:46:34 Pacific

Subject: Virus Help!!!!

101

FIG. 3

4/8

107



Sounds like you stopped it dead in its tracks. You could do a second virus scan at HOUSE CALL at the Trend Micro website if that would ease your mind. Brian, if you are going to use KAZAA, remember to be very cautious and scan evrything. As many people that try to help people at this security area know full well, there are virus/trojans/worms that are only available by using KAZAA!

**FIG. 4**

107

Sounds like you stopped it dead in its tracks. You could do a second virus scan at HOUSE CALL at the Trend Micro website if that would ease your mind. Brian, if you are going to use KAZAA, remember to be very cautious and scan <sup>111</sup> evrything. As many people that try to help people at this security area know full well, there are virus/ trojans/worms that are only available by using KAZAA!

FIG. 5

6  
G.  
E.

7/8

Computing.net Virus Bulletin (Microsoft), Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Search Favorites Help

Address: <http://www.computing.net/security/www/board/thread/4049.html>

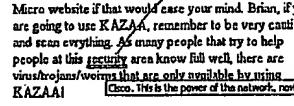
Please let me know - driving me crazy!

Thanks,  
Brian

**Report Offensive Message For Removal**

Response Number 1

Name: Capt  
Date: April 22, 2003 at 06:23:27 Pacific  
Subject: Virus Help!!!  
Reply:

Sounds like you stopped it dead in its tracks. You could do a second virus scan at **HOUSE CALL** at the Trend Micro website if that would ease your mind. Brian, if you are going to use KAZAA, remember to be very cautious and scan everything. As many people that try to help people at this **security** area know full well, there are virus/trojans/worms that are only available by using KAZAA! 

129

104

**Report Offensive Follow Up For Removal**

Response Number 2

Name: Brian  
Date: April 22, 2003 at 08:46:13 Pacific  
Subject: Virus Help!!!  
Reply:

thanks for your advice capt I am as careful as i can be on the catastrophe that is kuzza. And at least i know my av is working hah? lol

Thanks again for your time.  
Brian

**Report Offensive Follow Up For Removal**

Response Number 3

Name: Brian  
Date: April 22, 2003 at 08:46:24 Pacific  
Subject: Virus Help!!!

OPTIMIZE YOUR BUSINESS RELATIONSHIPS

133

101

<http://odserver.edtech.duq.edu/~k2j2z/003920/101.html>

FIG. 7

8/8

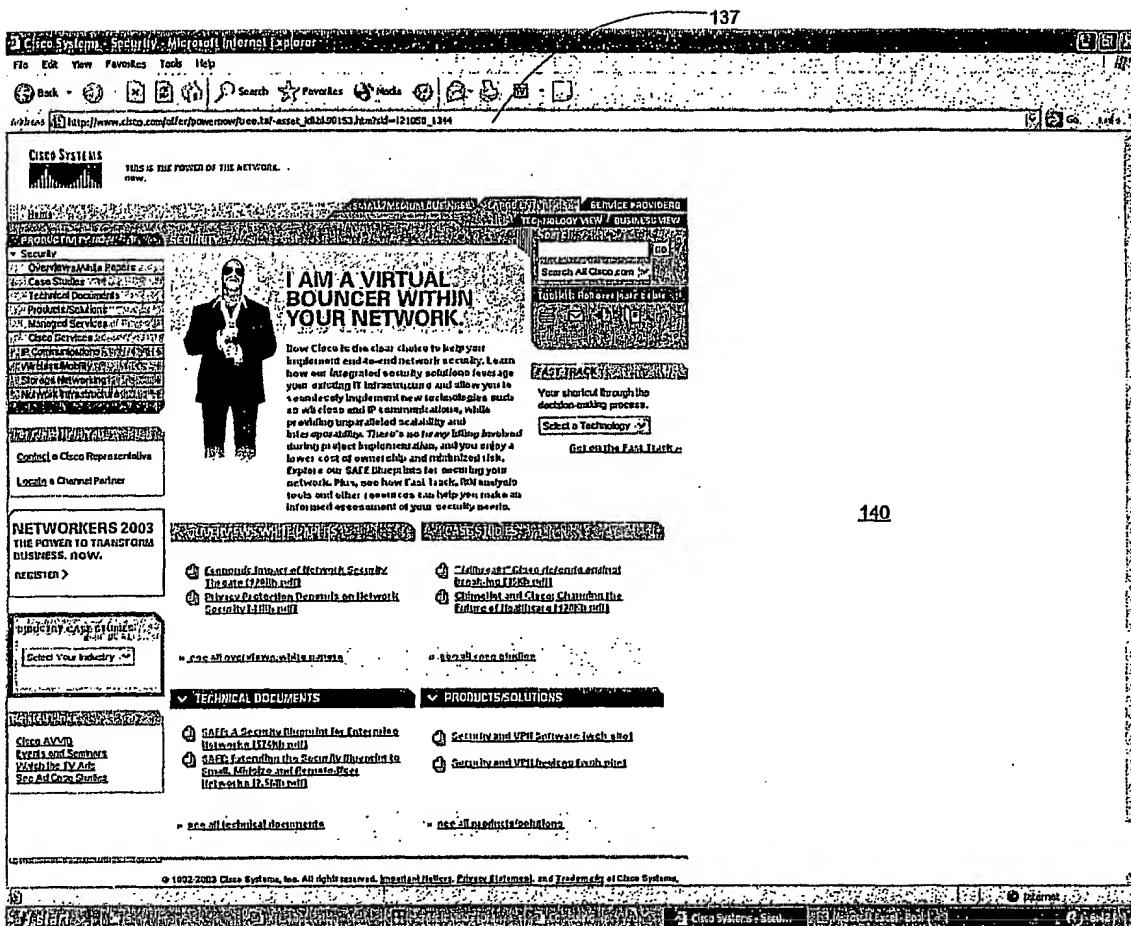


FIG. 8

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization International Bureau



(43) International Publication Date  
13 January 2005 (13.01.2005)

PCT

(10) International Publication Number  
WO 2005/003920 A3

(51) International Patent Classification<sup>7</sup>: G06F 17/30

(21) International Application Number: PCT/US2004/021294

(22) International Filing Date: 1 July 2004 (01.07.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/484,818 2 July 2003 (02.07.2003) US  
10/645,313 20 August 2003 (20.08.2003) US

(71) Applicant: VIBRANT MEDIA [US/US]; 19823 58th Place, South, Kent, WA 98032 (US).

(72) Inventors: STEVENSON, Douglas, B.; 4238 23rd Street, San Francisco, CA 94114 (US). GOODING, Craig, Allen; 30 Alma Street, London, GB NW5 (US).

(74) Agent: PATEL, Rajiv, P.; Fenwick & West LLP, Silicon Valley Center, 801 California Street, Mountain View, CA 94041 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:  
— with international search report

(88) Date of publication of the international search report:  
2 June 2005

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2005/003920 A3

(54) Title: METHOD AND SYSTEM FOR AUGMENTING WEB CONTENT

(57) Abstract: A system for augmenting data from a source data file with data from a reference database to generate an augmented data file is provided. The system includes a reference database including at least one reference datum. A handler component is configured to retrieve a source data file including the structured datum. A locator component is configured to locate the structured datum in the source data file; an analyzer component configured to associate the identified structured datum to one reference datum to create an association according to an analyzing strategy. A generating component is configured to generate a hyperlink based upon the association and embeds the generated hyperlink in the source file to create an augmented data file.

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US04/21294

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(7) : G06F 17/30

US CL : 707/005

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 707/005 709/219;706/52;704/7

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Continuation Sheet

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5,764,906 A (Edelstein et al.) 09 June 1998, see the entire document.	30-38,41-49,52-60, and 62
---		-----
Y		1-29,39-40,50- 51,61,63-82
Y	US 5,603,025 A (Tabb et al.) 11 February 1997, see the entire document.	1-15,19-29,63-77,79- 82
Y	US 6,098,065 A (Skillen et al.) 01 August 2000, see the entire document.	16- 18,39,40,50,51,61,78
A	US 6,578,078 B1 (Smith et al.) 10 June 2003.	1-82

Further documents are listed in the continuation of Box C.

See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier application or patent published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

14 February 2005 (14.02.2005)

Date of mailing of the international search report

04 MAR 2005

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US  
Commissioner for Patents  
P.O. Box 1450  
Alexandria, Virginia 22313-1450

Facsimile No. (703) 305-3230

Authorized officer

*7-1*  
BRIAN JOHNSON

*J. R. Johnson*  
Telephone No. (571) 272-3595

**INTERNATIONAL SEARCH REPORT**

International application No.  
PCT/US04/21294

Continuation of B. FIELDS SEARCHED Item 3:  
USPAT; US-PPGPUB  
program,software,record,text,character,keyword,fuzzy,expert,natural,engine,webpage

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record.**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**